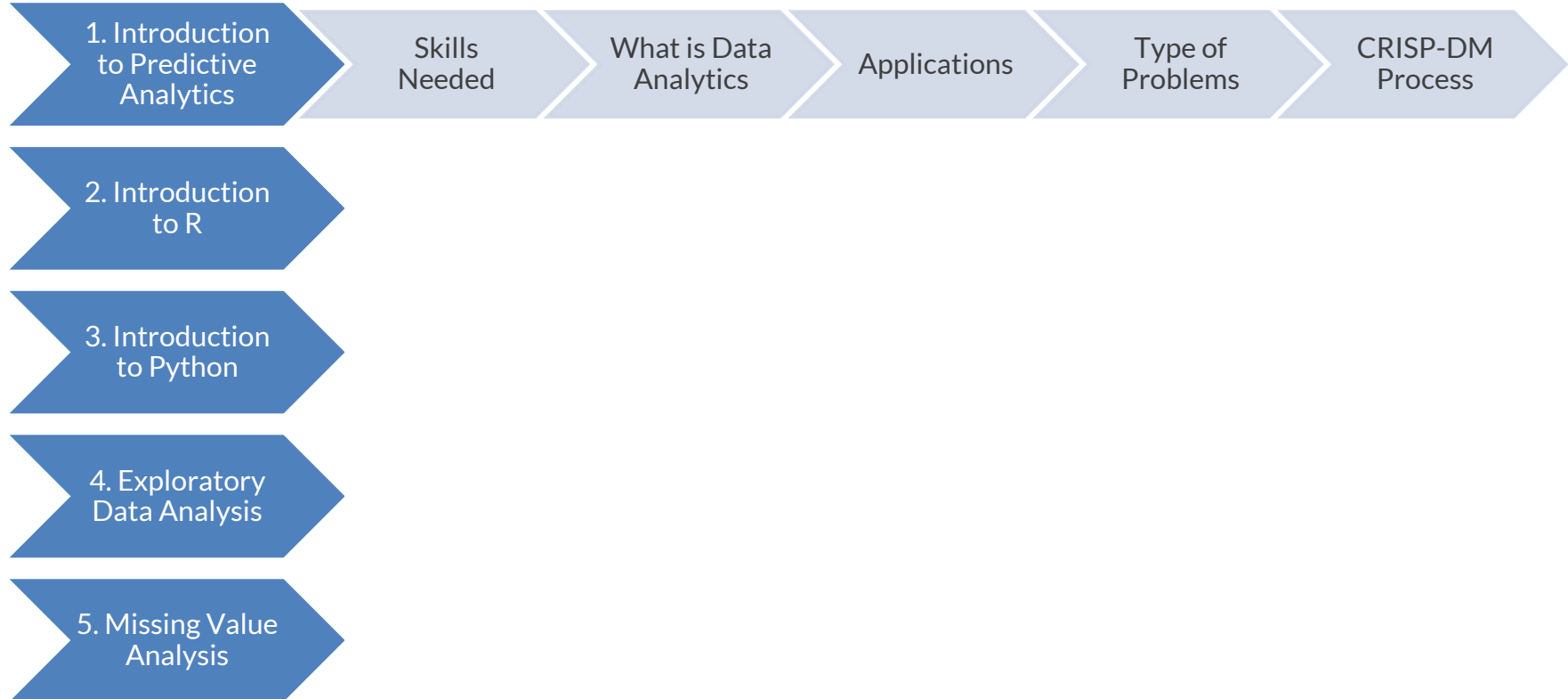




Predictive Analytics Using R and Python

- MUQUAYYAR AHMED
DATA SCIENTIST

We learnt!!



Today's Agenda

- Outlier Analysis

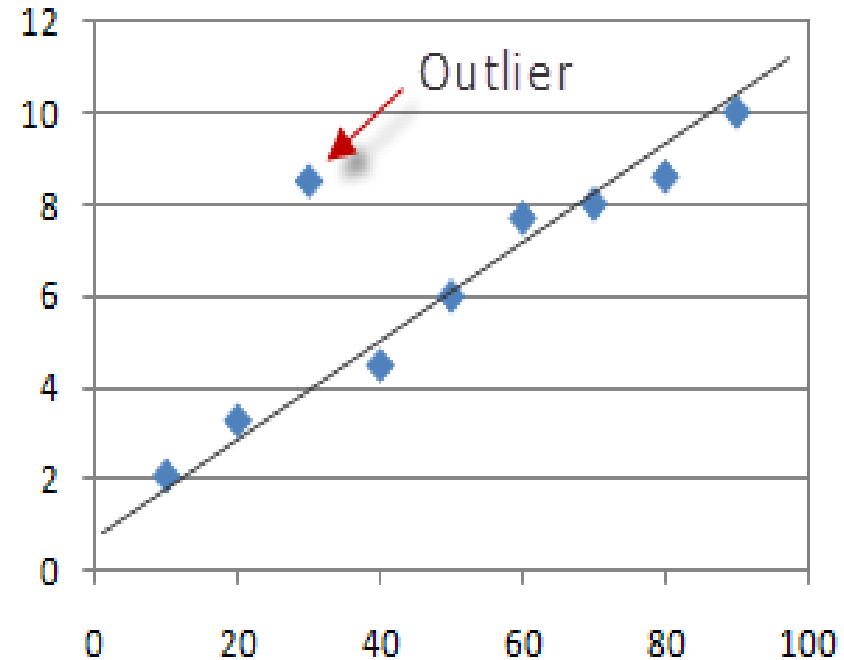
Outlier Analysis

What is an outlier?

- Observations inconsistent with rest of the dataset Global Outlier

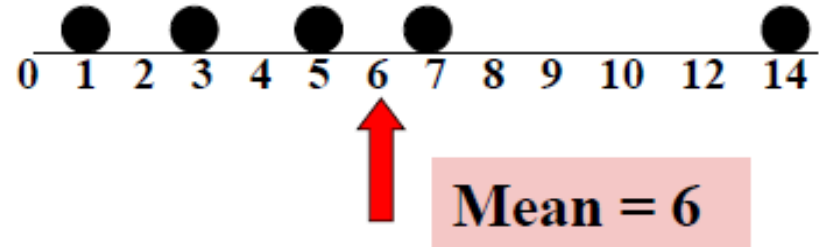
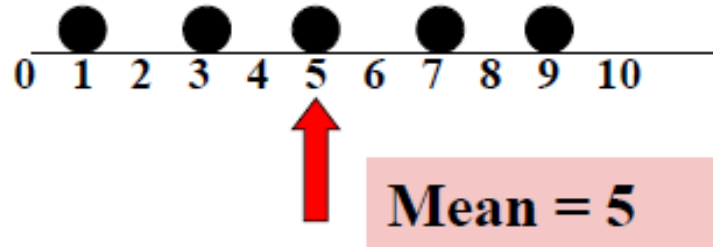
Causes of Outliers

- Poor data quality / contamination
- Low quality measurements, malfunctioning equipment, manual error
- Correct but exceptional data



Contd..

- Mean = Most Common measure of Central Tendency
- Mean Affected by Extreme Values



How to detect??

- Graphical tools
 - Box plot
- Statistical Technique
 - Grubbs' test for outlier
- R package "Outlier"
- Replace with NA

